DOCUMENT RESUME

ED 137 236 95 SP 010 873

AUTHOR Moore, Jeffrey E.; Berliner, David C.
TITLE The Maintenance of Data Security and the Trustworthiness of Individuals.
SPONS AGENCY National Inst. of Education (DHEW), Washington, D.C.
PUB DATE Apr 77
NOTE 13p.; Paper presented at the Annual Meeting of the American Educational Research Association, (New York, New York, April 4-8, 1977)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS Case Records; *Confidentiality; *Confidential Records; Data Collection; *Educational Research; Information Dissemination; *Legal Responsibility; Research Coordinating Units; Research Methodology; *Research Problems; *Statistical Data

ABSTRACT

This paper discusses the problem of maintaining confidentiality in data resulting from research projects. The rights and responsibilities of the parties involved in a research effort are examined. Five parties to a research project are identified as: (1) the participants, often called subjects, who have the right to privacy and the responsibility to provide honest, valid response; (2) the principal investigator, who selects or designs the collection effort and is responsible for maintaining privacy of subjects and providing accurate reports; (3) the funding agency, whose responsibilities often include selection of the principal investigator, approval of budget, auditing, and deciding whether or not to disseminate results; (4) the scientific community, usually in the form of a secondary analyst, auditor, or critiquer, whose rights include access to findings and data and whose responsibilities are to ensure that results are consistent with the data and analysis and to discourage improper practices and procedures; (5) the public, including the press and legislatures, who have a right to know the results of some research and in some cases have the responsibility for providing information to the general citizenry. The possible conflicts between the five identified parties in a research project are discussed as are means of clarifying roles of each; ways of insuring data security while disclosing important findings are examined. (JD)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED137236

24.17
F
SP

THE MAINTENANCE OF DATA SECURITY

AND THE

TRUSTWORTHINESS OF INDIVIDUALS

Jeffrey E. Moore

and

David C. Berliner

Far West Laboratory for Educational Research and Development

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

A paper presented at the meetings of the American Educational Research Association, New York City, New York, April 4-8, 1977.

The development of this paper was supported, in part, by the Beginning Teacher Evaluation Study, conducted for the California Commission for Teacher Preparation and Licensing, under funds provided by the National Institute of Education.

SP010 823

THE MAINTENANCE OF DATA SECURITY AND THE TRUSTWORTHINESS OF INDIVIDUALS

Jeffrey E. Moore and David C. Berliner
Far West Laboratory for Educational Research and Development

From almost any perspective issues of data security in educational research are multi-faceted, complex, and value laden. From our experience, these issues seem to revolve around conflicting perceptions of the rights and responsibilities of the parties involved in the research effort.

Five Parties to the Research Effort

The release of research data affects many parties, ranging from the individual or groups providing the basic data to the society which may have funded the study and hopes to feel some impact from the results. Each party in the research enterprise has some customary or traditional rights and responsibilities. Despite all the publicity when problems are aired in the press, only rarely have these rights and responsibilities been tested in court.

First there are the participants in the research study, often called subjects, but including some or all of the following: students, classes, schools, school districts, principals and teachers. The subjects certainly have the right to privacy, as well as the right to know the purpose for which the data will be used. A subjects' informed consent, which allows for responsible secondary analysis, should accomplish this. This broad type of consent is discussed by Dr. Winterbottom, in this symposium. The subjects' responsibility in a study is to provide honest and valid responses.

Another party in the research effort, and perhaps most central today, is the principal investigator for the research study. His rights include the right to select or design the collection effort, the analyses, and the reporting of the results. Responsibilities often include guaranteeing the privacy of the participants and ensuring that the results are reported accurately and with concern for the individuals whose consent and participation produces the raw data.

The next interested party is the funding agency, whose responsibilities ordinarily include the selection of the principal investigator, approval of the budget, auditing of the research effort and deciding whether or not to disseminate the results. One right of theirs, often in conflict with the principal investigator's, is the right to withhold data or summary reports of data from dissemination channels. Other rights include access to project personnel and material, accurate and timely reports on the progress of the project, and timely receipt of deliverables promised by the principal investigator.

The fourth interested party is the scientific community, usually in the form of a secondary analyst, auditor, or critiquer, whose rights include access to the findings of the research effort and access to the data for both verification and for additional or alternative analyses. Their responsibilities are to ensure that results reported are consistent with the data and analyses, to promote rigorous research and to discourage improper practices and procedures. The luxury of hindsight, afforded the secondary analyst, could easily lead to condemnation of the primary analysts' efforts. Thus a special responsibility is placed on the secondary analyst to be fair in judging the efforts of those who were working under more difficult conditions.

We also recognize a fifth party that sometimes is not well represented by the other parties to the research effort. This group is called the public and includes the press and legislature, as well as the general citizenry. These individuals all have a right to know, but often lack the sophistication to judge potentially generalizable research from site specific or non-replicable findings. The responsibility to provide information to that public has been conferred upon those who serve the public, e.g., reporters and legislators.

All of these parties coexist, usually without undue friction. When all interested parties have the same goals, respect one anothers' rights, and trust that each will carry out their responsibilities, there are few problems. However, since we are all here to discuss the issue of involuntary release of data, it is a sign that disagreement about each parties' goals, rights, and responsibilities is also quite possible. It is these conflicting perceptions among the parties to the research effort that we think are at the root of concerns about data security. Before proposing ways to alleviate some problems among the parties to the research effort we should identify what kinds of data are in need of protection against involuntary release, and what security systems exist for such data.

Types of Data

Last year representatives of the National Institute of Education, the California Commission for Teacher Preparation and Licensing, its' Advisory Board, Educational Testing Service, and the Far West Laboratory for Educational Research and Development discussed issues involved in the release of data from a large, multi-year, multi-contractor, multi-million dollar educational research study titled the Beginning Teacher

Evaluation Study. The meetings resulted in the categorization of data into two major groups. First, there is "sensitive" data which may be defined as data which would permit the identification of individual participants (students, schools, districts, etc.) in a study. It is "sensitive" when such individual identification could lead to public embarassment or ridicule of the participant in the research study. Examples of "sensitive" data include student test booklets with names attached, memos which include teacher names in a study of teacher effectiveness, lists of schools to be visited, and certain combinations of demographic data (i.e., sex, SES, age, teacher, and school) which could permit easy identification of a participant in the study. Admittedly, this last example is not as clear an example of "sensitive" data as the one preceeding it, but there is a substantial "gray" area which prevents simple solutions to the problems of data security. The "sensitive" data might very well be of such a nature that it is to be destroyed as soon as possible, or, at least, guarded dilligently. The research staff might be ready, perhaps, to go to jail to prevent disclosure to anyone not authorized to examine the data.

The second type of data is referred to as "archival" data, defined as data which does not permit "easy" identification of individual participants. Examples of "archival" data are test booklets with only an I.D. number encoded, memos and lists which refer to participants by number only, and "limited" combinations of demographic data. Of course with "archival" data there are usually many sets of "keys" which have both names and numbers, and these lists must be classified as "sensitive." For the archival data we have few qualms about its release to other investigators, funding agencies, or the public in general.

Data Security

Historically, in educational research, the data collected by a researcher was directly under his or her control. From collection, through analysis, to the final writing of reports, the security of the data was the responsibility of the principal investigator. Recent developments, including intensive examinations of social science methodology, increased calls for secondary analyses, increased use of large data base systems, remote computer access and investigative reporting have increased the desire of many individuals for access to research information. In our opinion, there has not been nearly enough consideration given to the physical security of data as changes in modern social science research have occurred and desire for access to data has multiplied. The present security systems, though involving computers, are for the most part, not very different from the systems that were in force when data was kept in file folders and nobody ever sued for invasion of privacy. The typical educational researcher is usually a victim of naivete, supporting a belief in the inherent goodness of his fellow man. In addition, he or she is probably suffering from a lack of funds and a lack of time, which leads to carelessness in the maintenance of security for data. For these reasons we are bound, we think, to have some rather juicy lawsuits and scandals emerge in the next decade as social scientists in education learn what they should and should not do in these rapidly changing times.

The above opinions about the current state of affairs are also an indictment of our own data security procedures. We exercise what we call "reasonable care." But it is clear to us that anyone who really sets out to steal our most sensitive data, to embarrass us and the research participants publicly, can do so regardless of the greatest expenditures we may

make. Even the tightest security systems can be broken by determined people. Therefore, in our opinion, concern about the involuntary release of data should reflect concern for the roles of people involved in the research enterprise. Thus we will not discuss any of the sophisticated procedures that may be taken to provide for physical security in order to prevent the involuntary release of data. Rather, since people are the reasons for any difficulties with data security, we intend to discuss how people may work together to optimize the relationship between those with a desire for access to data and those responsible for the voluntary release of the data.

Dissemination of Data

Because of the potential sensitivity of virtually any aspect of a study, it appears that the most cautious of principal investigators could end up trying to protect most of their data. However, at least in principle, investigators should be able to identify some of their data as archival and should agree also that it can be released to certain parties without severe restrictions on the use of that data. The principal investigator may also specify that other data, patently more sensitive data, should not be released except under very special conditions. Problems with this principle occur because a good deal of data is in a kind of a "twilight zone": data which have no names attached, only numbers, and for which some demographic information is available. A whole set of questions are raised about such data. Should the data be released to anyone who wants it or should the data be released only to special people? Who decides the trustworthiness of those requesting data? What are the requesters of data going to do with the data, for example, will they give the data to someone else? Who will provide the resources for the tasks needed in preparing and releasing

data? These questions are important, usually not discussed at the beginning of a research study, and very much in need of consideration.

The individuals and parties to the research effort find out the extent of their mutual trust in each other and the degree of shared perceptions, when questions about the sensitivity of data and their use are raised. Historically, the key person to decide which data was to be released, and to whom it should go, has been the principal investigator. In lieu of any special agreements, we think the principal investigator still must retain that decision-making authority. This position is not taken lightly. It is the principal investigator who, customarily, has made the guarantee of privacy to the participants. We do not argue that this state of affairs must necessarily continue, but such is the typical state of affairs. Certain problems are inherent in this situation. For example, how long is an individual researcher to maintain control over release of the data? May he transfer responsibility to other investigators within or outside the project, or, like the captain of a ship, will he be held responsible for a collision even if he wasn't on the bridge? Is the investigator obligated to provide copies of the raw data tapes to the funding agency, to the scientific community, or to the general public? Is the principal investigator to pay the costs for releasing data even after the project has ended?

Where does one party's responsibilities infringe upon the rights or responsibilities of another party? Does the party which pays the bills have a right to examine the data or to have the names of participants in order to ensure that they were in fact participants. There are many more questions like these which could be raised and situations which could be posed for all data considered to be sensitive by the principal investigator. How can we begin to answer them all?

In our opinion, the issues can be settled, reasonably well, if we admit that the trust which is characteristic of the parties at the start of a research study is not enough. Just as a handshake may not suffice to protect the interested parties in a complex business transaction, unspecified trust may not be enough to protect all interested parties in a complex large scale educational research study. Therefore, we make the following suggestions in the interest of all the concerned parties.

First, decide early (preferably before a Request for Proposal is written) whether the funding agency or the principal investigator will guarantee protection of privacy to the participants in the study. We suggest that, in the absence of any written agreement, the implied guarantor is the party which asks for the informed consent of the participants in the study. And therefore, this party should have the final word in decisions regarding the release of any data collected in the course of the study. Ordinarily, this party is the principal investigator, but it need not be.

Second, decide early what degree of publicity is to attend the final report. To protect the interests of the public, the scientific community, and the principal investigator and project staff, we feel it is imperative to know whether the results of the proposed study are to be widely disseminated, or is there the possibility that the results may be held back? More than once it has not been to the best advantage of a funding agency to have negative results widely publicized. Does the agency have the right to stop the dessemination of reports and data by the principal investigator? We believe that the degree of publicity, like the level-of-significance, should be decided <u>a priori</u>. Such actions could solve the potential for much conflict in this area. The degree of documentation for the data files

will be known if the degree of publicity for the research effort is thought about in advance. For example, if it is believed that press conferences will be used to announce the findings of a national study of special educational programs for low income children, the principal investigator will take a set of actions different than if he believed that a committee of school board members are the audience for a report on the effects of increased per pupil expenditures on student achievement.

Following from the preceding suggestions are our recommendations regarding volunatry release of data for secondary analysis. If the results of the study are to be publicized, archival data should be released to any interested party willing to pay reasonable costs for the data. Potentially sensitive data may be released to those parties which are able to: a) rationalize their need-to-know through a written proposal; b) provide a written guarantee to maintain security of the data: and, c) pay the extra costs associated with preparing the specially edited tape required for their purposes. Nevertheless, the final decision as to whether any sensitive data will be released must remain with the party which has guaranteed privacy to the participants. This power and its attendant responsibilities are not assignable. The final responsibility for the maintenance of data security is the guarantor's, and will remain so unless released from this obligation by the participants. In terms of the BTES study, to the possible distaste of the Commission, the National Institute of Education, and interested secondary analysts, we say that we, not they, must decide what sensitive data will be released and to whom it will go since we insured the participants' rights to privacy. Hopefully the trust that exists between the parties to this research study will continue. But with a little more attention to these matters at the beginning of a study,

such trust can be made the subject of explicit contractual obligations, relieving all parties from the problems inherent in ambiguity.

Our last suggestion deals with costs, mentioned above, of preparing data for secondary analyses. If secondary analyses of the data are envisioned when the study is planned, the funding agency must see that sufficient project resources have been allocated to provide for the preparation of archival and other master data tapes and accompanying documentation. It is our experience that under the pressures of meeting deadlines for final reports, while trying to stay within the budget, only those data immediately needed and only that minimal documentation required by the staff are prepared. We recommend that a separate work unit or task be required which details the data to be kept for secondary analysis and the level of documentation to be provided with the data. We believe that implementation of this suggestion will cost less in the long run than the present haphazard methods. The necessary data and documentation may be prepared as the project progresses, rather than having the project managers go through a scramble at the end of the project year when key personnel may have left, along with the working versions of documentation and/or, heaven forbid, the data.

Summary

In conclusion, we restate that in lieu of some contractual specifications, the principal investigator retains the responsibility for deciding how sensitive data will be released and to whom. Trustworthiness among the parties in the research enterprise is desirable, but such trust among individuals is considerably enhanced when certain commitments are made clear, in advance, in the RFP or the contract. Contracts of this type

may circumscribe a principal investigator's traditional rights. This may be agreeable to some and unconscionable to others. We simply think that the nature of the trustworthiness and wisdom of the principal investigator need not be left open-ended. Rather, the limits of the rights and responsibilities of the principal investigator can be spelled out in the contract for the project. With forethought and planning, the rights of all interested parties to the research study may be protected, money saved, and primary data made available for secondary analyses.